



University  
of Glasgow

Fleming, J.R., Morgan, R.E., Fyfe, P.K., Kelly, S.M., and Hunter, W. (2013) *The architecture of trypanosoma brucei tubulin-binding cofactor B and implications for function*. FEBS Journal, 280 (14). pp. 3270-3280. ISSN 1742-464X

Copyright © 2013 The Authors

<http://eprints.gla.ac.uk/80282/>

Deposited on: 04 March 2014

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>

# The architecture of *Trypanosoma brucei* tubulin-binding cofactor B and implications for function

Jennifer R. Fleming<sup>1</sup>, Rachel E. Morgan<sup>1</sup>, Paul K. Fyfe<sup>1</sup>, Sharon M. Kelly<sup>2</sup> and William N. Hunter<sup>1</sup>

<sup>1</sup> Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, UK

<sup>2</sup> Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, UK

## Keywords

CAP-Gly domain; CD; crystallography; tubulin-binding; ubiquitin-like

## Correspondence

W. N. Hunter, Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK  
Fax: +44 1382 385764  
Tel: +44 1382 385745  
E-mail: w.n.hunter@dundee.ac.uk

(Received 4 February 2013, revised 1 April 2013, accepted 23 April 2013)

doi:10.1111/febs.12308

Tubulin-binding cofactor (TBC)-B is implicated in the presentation of  $\alpha$ -tubulin ready to polymerize, and at the correct levels to form microtubules. Bioinformatics analyses, including secondary structure prediction, CD, and crystallography, were combined to characterize the molecular architecture of *Trypanosoma brucei* TBC-B. An efficient recombinant expression system was prepared, material-purified, and characterized by CD. Extensive crystallization screening, allied with the use of limited proteolysis, led to structures of the N-terminal ubiquitin-like and C-terminal cytoskeleton-associated protein with glycine-rich segment domains at 2.35-Å and 1.6-Å resolution, respectively. These are compact globular domains that appear to be linked by a flexible segment. The ubiquitin-like domain contains two lysines that are spatially conserved with residues known to participate in ubiquitylation, and so may represent a module that, through covalent attachment, regulates the signalling and/or protein degradation associated with the control of microtubule assembly, catastrophe, or function. The TBC-B C-terminal cytoskeleton-associated protein with glycine-rich segment domain, a known tubulin-binding structure, is the only such domain encoded by the *T. brucei* genome. Interestingly, in the crystal structure, the peptide-binding groove of this domain forms intermolecular contacts with the C-terminus of a symmetry-related molecule, an association that may mimic interactions with the C-terminus of  $\alpha$ -tubulin or other physiologically relevant partners. The interaction of TBC-B with the  $\alpha$ -tubulin C-terminus may, in particular, protect from post-translational modifications, or simply assist in the shepherding of the protein into polymerization.

## Introduction

Key to the regulation of many biological processes is the tight control exercised over protein biosynthesis, folding, and degradation. With respect to tubulin, a central component of the cytoskeleton, the correct folding and polymerization involves distinct stages that are influenced by several chaperones or cofactors [1,2]. Initially, after a tubulin polypeptide is produced, it is

captured by prefoldin [3] and subsequently passed to chaperonin-containing T-complex polypeptide 1 (CCT) [4]. When released from CCT, tubulin is essentially folded, but appears to be unable to polymerize and form microtubules (MTs) [1]. Our understanding of what occurs between release from CCT and MT formation is limited. Five tubulin-binding cofactors

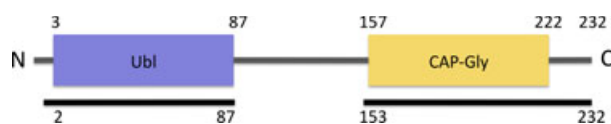
## Abbreviations

CAP-Gly, cytoskeleton-associated protein with glycine-rich segment; CCT, chaperonin containing T-complex polypeptide 1; CLIP-170, cytoskeleton-associated protein with glycine-rich segment domain containing linker protein 170; ITC, isothermal titration calorimetry; MT, microtubule; PDB, Protein Data Bank; SeMet, selenomethionine; TBC, tubulin-binding cofactor; TbTBC-B, *Trypanosoma brucei* tubulin-binding cofactor B; TEV, tobacco etch virus; Ubl, ubiquitin-like.

(TBCs) are implicated in late-stage tubulin folding and heterodimer assembly [1,5–10]. TBC-B and TBC-E are implicated in binding  $\alpha$ -tubulin, whereas TBC-A and TBC-D interact with  $\beta$ -tubulin. TBC-C is involved in the final stages of dimer formation, stimulating GTP hydrolysis in  $\beta$ -tubulin and heterodimer release from a protein assembly [2]. Little is known about the molecular basis for the roles of these cofactors, a point that we sought to address in relation to TBC-B.

TBC-B comprises an N-terminal ubiquitin-like (Ubl) domain and a C-terminal cytoskeleton-associated protein with glycine-rich segment (CAP-Gly) domain. There are NMR structures of the TBC-B Ubl domains from *Caenorhabditis elegans* [11], *Drosophila melanogaster* [Protein Data bank (PDB) code [2KJR](#)], *Arabidopsis thaliana* (2KJ6), and *Mus musculus* (1V6E). The mouse TBC-B CAP-Gly domain NMR structure has also been determined (1WHG), but the only crystal structure known is that of the *C. elegans* TBC-B CAP-Gly domain [12]. We targeted the protein from *Trypanosoma brucei*, an organism that is considered to be a useful model for the study of MT biology [13]. Searching against the translated *T. brucei* genome (<http://www.genedb.org/>) with mouse, human and *C. elegans* TBC-B sequences identified a single protein, Tb10.61.2930, with ~40% amino acid sequence identity. This protein, *T. brucei* TBC-B (*TbTBC-B*), consists of 232 amino acids organized into the N-terminal Ubl domain (~90 residues) and a CAP-Gly domain (residues 157–222; Fig. 1).

We now report the construction of an efficient bacterial recombinant expression system, protein purification and the use of CD and bioinformatics approaches to investigate secondary structure content and predicted flexibility. Crystallization of the full-length protein was achieved, but the samples were poorly ordered. Consequently, structural analyses of the individual domains were carried out. The fortuitous observation in the crystal structure of the CAP-Gly domain of intermolecular contacts with the C-terminus of a symmetry-related molecule suggests how interactions with partners, including  $\alpha$ -tubulin, might occur.



**Fig. 1.** Schematic of *TbTBC-B*. Domains and termini are labelled. Residues 3–87 and 157–222 represent the assigned Ubl (blue box) and CAP-Gly (yellow box) domains, respectively. Black lines and residue numbers 2–87 and 153–232 represent the structures determined.

Comparisons with ubiquitin suggest a functional link between the structure of TBC-B and the regulation of distinct populations of proteins associated with tubulin biology by proteasome-dependent degradation.

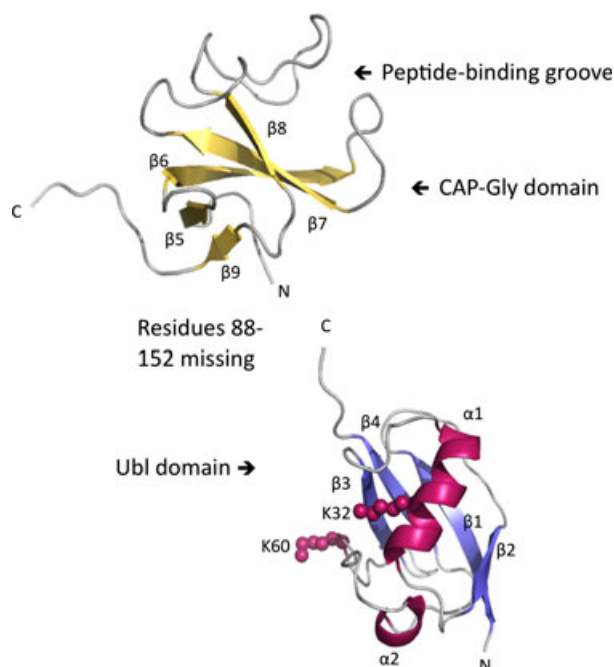
## Results and Discussion

Recombinant full-length *TbTBC-B* was produced in *Escherichia coli* and purified. The final step in purification was size exclusion gel chromatography, and this indicated that the full-length protein is monomeric in solution. The protein crystallized readily; however, despite displaying a good appearance (data not shown), X-ray diffraction from the crystals did not extend beyond 7 Å of resolution. Limited proteolysis produced two polypeptides, one of which, the Ubl domain, gave highly ordered crystals, and the structure was solved with anomalous dispersion methods [14]. Subsequently, crystals of the CAP-Gly domain were also obtained after testing of a series of truncated constructs and molecular replacement allowed for structure solution. We note, that following proteolysis, the two domains were readily separated and purified, indicating that there was no domain–domain association, and when studied individually they remained monomeric in solution.

### The Ubl domain

The N-terminal Ubl domain of *TbTBC-B* crystallized in the tetragonal space group  $P4_12_12$  with a single polypeptide in the asymmetric unit. Size exclusion chromatography employed during purification indicated that the protein was monomeric in solution.

The Ubl domain is a small globular entity consisting of a mixed four-strand  $\beta$ -sheet that forms a concave groove in which a single  $\alpha$ -helix is placed (Fig. 2). This is an example of a  $\beta$ -grasp fold, a common structure involved in protein–protein interactions [5,11]. A pronounced hydrophobic core is present, formed mainly by aliphatic side chains. The residues involved, which have <10% solvent-accessible surface area, include Val4, Val6, Leu8, Tyr22, Ile28, Ile31, Val35, Thr41, Met46, Leu48, Leu50, Met62, Leu68, Cys73, Ile79, and Val81 (not shown). These core residues are, in general, conserved among the TBC-B family of proteins. The surface residues are more variable, perhaps indicative of a module that does not interact directly with  $\alpha$ -tubulin. Note that, as  $\alpha$ -tubulin is a very highly conserved protein, any interacting components would be expected to display conservation on their surfaces if they interacted at similar positions.



**Fig. 2.** Ribbon diagram of the two domains of *TbTBC-B*. The Ubl domain  $\beta$ -strands are in light blue,  $\alpha$ -helices are in pink, and the CAP-Gly domain  $\beta$ -strands are in yellow. The N-termini and C-termini and elements of secondary structure are labelled. The orientation of the two domains with respect to each other is arbitrary. Two lysines of note are depicted as pink spheres.

The Ubl module, in terms of amino acid sequence, is the more variable domain of TBC-B (Fig. 3) as compared with homologues. For example, the Ubl and CAP-Gly domains of *T. brucei* and mouse TBC-Bs share ~30% and 55% sequence identity, respectively. Although remarkably similar in structure to ubiquitin (rmsd of 0.83 Å over 41 C $\alpha$  atoms with PDB code [1UBQ](#)), the *TbTBC-B* Ubl domain shares only 10% sequence identity. Ubiquitin contains seven lysines (residues 6, 11, 27, 29, 33, 48, and 63) that are targets for covalent modification [15,16]. Intriguingly, and despite only low sequence conservation, the Ubl

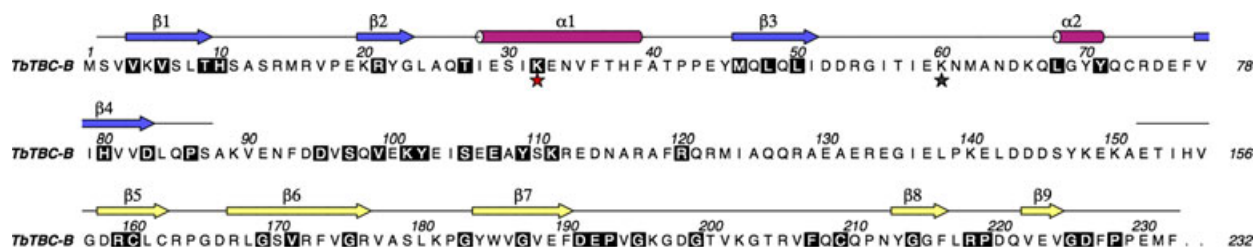
domain of *TbTBC-B* contains two lysines (Lys34 and Lys62), which correspond to Lys27 and Lys48 of ubiquitin, and a structural overlay reveals strong structural similarity in terms of the positioning of these residues (Fig. 4). The Ubl domains of known structure overlap at a similar level (Fig. 5A).

### The CAP-Gly domain

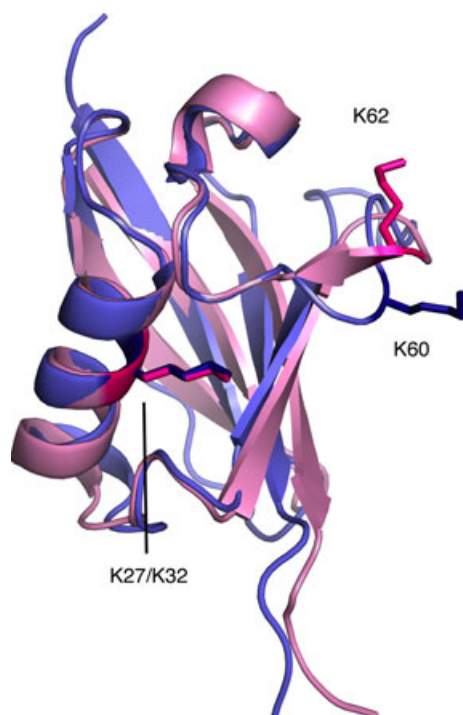
Crystals of the CAP-Gly domain of *TbTBC-B* were obtained from a construct producing the 81 C-terminal amino acids. The peptide eluted from the size exclusion column as a single species of ~11 kDa, indicating a monomer in solution. The crystals are orthorhombic with space group  $P2_12_12_1$ , with two polypeptides, labelled A and B, in the asymmetric unit. The rmsd observed after superposition of 79 C $\alpha$  positions of molecule A on molecule B was 0.86 Å. No restraints were imposed on noncrystallographic symmetry during refinement, so this indicates that the molecules are highly similar. The numbering of residues and secondary structure elements in the C-terminal domain is carried on from the Ubl domain.

Although the sequence derived from the genome assigns residue 223 as aspartate, the cloned gene encodes a glutamate at this position, an observation confirmed by the structural analysis. This conservative difference could be an artefact of the cloning process, or, more likely, could result from a natural variation in the strain that from which genomic DNA was obtained. We note that the *T. brucei gambiense* gene sequence for TBC-B (<http://www.genedb.org/>) encodes a glutamate in this position.

The *TbTBC-B* CAP-Gly domain is a small globular structure with a similar fold to other CAP-Gly domains, and is an important module for the recognition and binding of the C-terminal tail of  $\alpha$ -tubulin [17,18]. Structural comparison reveals that *TbTBC-B* matches closely to other CAP-Gly domains (Fig. 5B). The *C. elegans* TBC-B CAP-Gly domain, which shares ~50% sequence identity, is the closest struc-



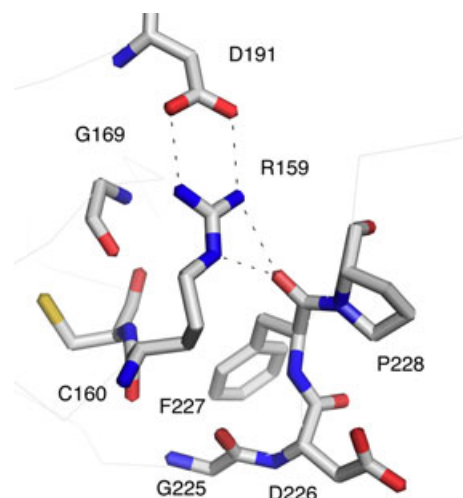
**Fig. 3.** The primary and secondary structure of *TbTBC-B*. Residues strictly conserved in at least 75% of TBC-B sequences are enclosed in black. Red stars mark Lys32 and Lys60, which are conserved in ubiquitin.



**Fig. 4.** Structural overlay of *TbTBC-B* Ubl (blue) and ubiquitin (PDB code [1UBO](#); pink). Spatially conserved lysines are depicted as sticks.

tural relative, with an rmsd of 1.4 Å for 95 C $\alpha$  atoms. The fold is dominated by five  $\beta$ -strands,  $\beta$ 5– $\beta$ 9, forming a consecutive twisted antiparallel sheet with  $\beta$ 9 returning to lie in an antiparallel fashion at the other side of  $\beta$ 5. The side of strand  $\beta$ 8 forms the floor of a solvent-exposed, primarily basic groove, which is flanked by the two extended loops, linking  $\beta$ 6 and  $\beta$ 7, and  $\beta$ 7 and  $\beta$ 8. NMR studies have revealed that the  $\alpha$ -tubulin C-terminal tail peptide binds to this groove in the CAP-Gly domain of human CAP-Gly domain containing linker protein 170 (CLIP-170) [19]. Several glycines (residues 170, 188, 195, 200, 215, and 225), which are responsible for the name of this domain, are highly conserved and contribute to the fold of the domain [17]. Residues around the peptide-binding groove are also highly conserved, Phe216, Leu180, Trp185, Val201 and Phe207 forming a hydrophobic core that helps to stabilize the floor of the groove.

On the other side of the domain from the peptide-binding groove, there are two salt bridges formed by Arg159 with Asp191, and Arg172 with Glu189, which link  $\beta$ 5 with  $\beta$ 7, and  $\beta$ 6 with  $\beta$ 7, respectively. In addition, hydrogen bonds between Arg159 and the backbone oxygen of Phe227 link  $\beta$ 5 to the C-terminal segment of  $\beta$ 9 (Fig. 6).

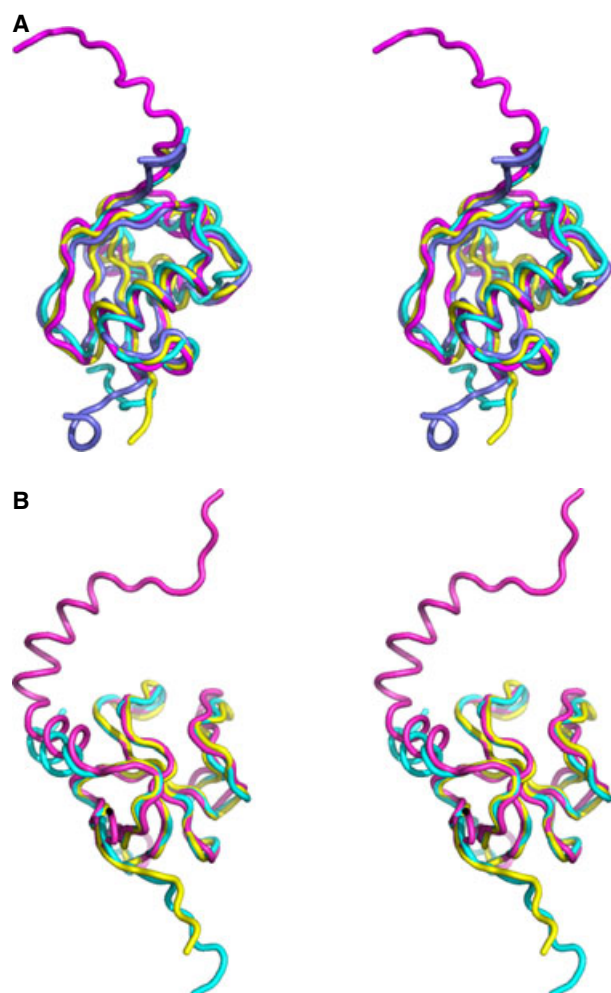


**Fig. 5.** (A) The *TbTBC-B* Ubl domain superimposed with other TBC-B Ubl structures. Structures are shown as tubes coloured by structure: yellow, *TbTBC-B*; cyan, *C. elegans*; magenta, *D. melanogaster*; slate, *M. musculus*. Affinity tags, where present, have been omitted from the models. The N-terminus is at the top, and the C-terminus points to the bottom. (B) The *TbTBC-B* CAP-Gly domain superimposed on other TBC-B CAP-Gly domains, in a similar way as in (A): yellow, *TbTBC-B*; cyan, *C. elegans*; pink, *M. musculus*. The N-terminus is on the left side, and the C-terminus is on the right.

CAP-Gly domains, including those of TBC-B, possess a highly conserved pentapeptide tubulin-binding motif with a sequence that is almost always GKNDG [19,20]. This motif is placed on one of the loops that flank the peptide-binding groove. NMR studies of CLIP-170 indicate that the asparagine at position 3 can participate in hydrogen bond formation with the C-terminal tail of peptides ending with the sequence EE[Y/F]. This is the sequence found at the C-terminus of  $\alpha$ -tubulin and several MT tip-binding proteins. Occasionally, the asparagine in the tubulin-binding motif is replaced by a histidine, a conservative change that does not affect peptide binding. However, mutating either the lysine or the asparagine to alanine has a deleterious effect on binding [20]. The *TbTBC-B* CAP-Gly domain has a different sequence in this tubulin-binding motif at residues 195–199, GKGDG. Glycine rather than asparagine occupies the third position, and this nonconservative change precludes the formation of a side chain hydrogen bond with the terminal residue of an EE[Y/F] peptide.

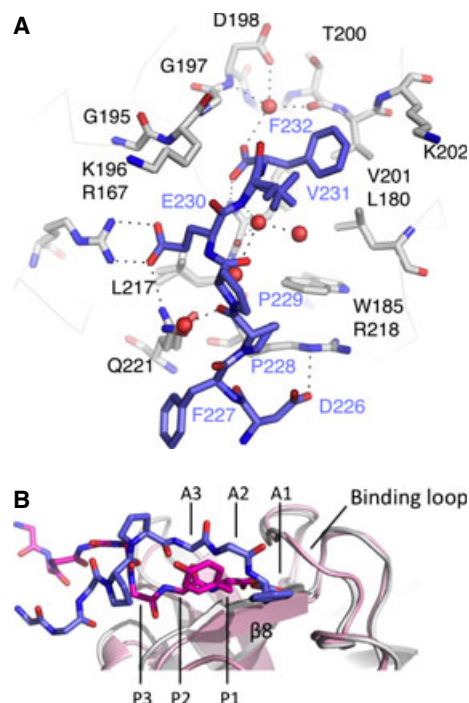
No crystal structures have been solved with an  $\alpha$ -tubulin tail-like ligand bound to TBC-B, and, despite our efforts, it was not possible to cocrystallize the *TbTBC-B* CAP-Gly domain with peptides either. However, and fortuitously, in our structure the C-terminus





**Fig. 6.** The Asp191 and Arg159 salt bridge in the CAP-Gly domain. The main  $\alpha$  trace is shown as grey sticks. Hydrogen bonds are shown as black dashed lines. Conserved residues are shown as sticks coloured by atom type: C, grey; N, blue; O, red, S, yellow.

of chain A forms intermolecular contacts with the basic groove of a symmetry-related chain B (Fig. 7A). The symmetry operation is  $(-x + 1/2, -y, z + 1/2)$ . The loss of these intermolecular contacts may help to explain why constructs in which the C-terminus had been truncated did not crystallize. The C-terminus of *TbTBC-B* has the sequence EVF, which is similar to the  $\alpha$ -tubulin tail EE[Y/F] motif, and so these intermolecular interactions can be taken to mimic the association with a relevant binding partner. The terminal residue of chain A, Phe232, anchors the peptide in the groove with a combination of hydrophobic and hydrophilic interactions. The aromatic side chain interacts with a hydrophobic patch consisting of the chain B residues Leu180, Val201, and Phe216. The carboxylate



**Fig. 7.** (A) Chain A C-terminus interacting with the peptide-binding groove of a symmetry-related molecule, chain B. The chain B main chain is shown in grey ribbon style, and residues that surround the C-terminal tail are shown as sticks coloured according to atom type: C, grey; N, blue; O, red. Chain A is shown as sticks coloured according to atom type: C, slate blue; O, red. Hydrogen bonds are shown as black dashed lines. Residue numbers for chain A are shown in cyan. L217 and Q221 (chain B) and V231 (chain A) are shown with two side chain rotamer conformations. G199 and F216 are not labelled. (B) Relative positions of the C-terminus peptides in *M. musculus* CAP-Gly domain II of CLIP-170 structure (pale pink cartoon; peptide carbon atoms are in magenta) and *TbTBC-B* (chain B as a white cartoon; chain A carbon atoms are in slate). For both peptides: O, red; N, dark blue. Amino acids, for clarity, are shown as backbones, except for proline and the C-terminus residues. Peptides are labelled 1–3 from the C-terminus: A for *TbTBC-B* chain A; P for *M. musculus* polypeptide.

of this terminal residue forms a hydrogen bond with the main chain amide of Phe216, and participates in a water-mediated hydrogen-bonding network with Asp198 and Thr200. The C-terminal peptide then arches away from  $\beta 8$ , trailing back to Glu230, which forms a salt bridge with Arg167. Another residue, Gln221, interacts with Glu230, as well as contributing to a water-mediated interaction with the backbone of Pro228. Gln221 is conserved in the trypanosomatid TBC-B sequences, but is highly variable in other species. The final interaction that chain A undergoes as it exits the groove is a salt bridge between Asp226 and Arg218.

The NMR structure of the CAP-Gly domain of human CLIP-170 complexed with an  $\alpha$ -tubulin tail peptide provides an example for comparative purposes [20]. The domains share  $\sim 50\%$  identity and are structurally well conserved, with an rmsd of 1.06 Å over 50 C $\alpha$  atoms. In the CLIP-170  $\alpha$ -tubulin tail peptide complex, the peptide is positioned closer to the  $\beta$ 8 strand (Fig. 7B).

The Gly/Asp difference at position 4 in the tubulin-binding loop, discussed earlier, allows the C-terminus of chain A to bind further along the groove than the  $\alpha$ -tubulin tail peptide in the CLIP-170 structure. This causes a relative shift in the position of the Val231 side chain, which is directed out of the groove, and undergoes no interactions with the CAP-Gly domain. The third residue from the end of chain A, Glu230 (A3 in Fig. 7), is in the same position as the second residue from the end in the  $\alpha$ -tubulin tail peptide, Glu450 (P2 in Fig. 7). This may explain why the sequence EVF can bind, contrary to the conclusion that EE[Y/F] is the essential recognition sequence [18,19]. In an attempt to further investigate CAP-Gly peptide interactions, we carried out isothermal titration calorimetry (ITC) with the hexapeptide EDVEEY, which represents the C-terminal residues of *T. brucei*  $\alpha$ -tubulin. A range of concentrations of the protein domain and the peptide were tested, but no heat changes were observed (data not shown).

The peptide-binding groove on chain A is occupied by two formate molecules, derived from the crystallization mixture, which bind in similar fashion to the negatively charged moieties of the C-terminal peptide just discussed (data not shown). One formate interacts with Arg167 and the other with Gln221.

### Unique features of *Trypanosoma* TBC-B

CAP-Gly domains recognize the highly conserved  $\alpha$ -tubulin C-terminal tail with the sequence EE[Y/F] [18]. In *T. cruzi* and *Trypanosoma vivax*  $\alpha$ -tubulin, this sequence is matched exactly. In *T. brucei gambiense* and *T. brucei* 427, the sequence is EMF, and, as just described, we present structural data to show that the sequence EVF can also bind to a CAP-Gly domain. Our structure of *Tb*TBC-B indicates that the penultimate residue, whether glutamate or methionine, is probably directed out of the peptide-binding groove, and, with no direct interactions involving the side chain, the identity would appear to be less important for binding.

Both TBC-B and TBC-E retain highly conserved CAP-Gly domains across numerous different species, and this probably reflects important roles in tubulin

biology [20]. The CAP-Gly domain of TBC-E also usually contains the GKNDG tubulin-binding motif. The closest homologue to human TBC-E in *T. brucei* is Tb927.3.2680, a 530-residue protein with  $\sim 25\%$  sequence identity. The protozoan protein contains a leucine-rich repeat segment and a Ubl domain but, surprisingly, lacks a CAP-Gly domain. In fission yeast, the TBC-E homologue Alp21 also lacks a CAP-Gly domain, but is indispensable for maintaining  $\alpha$ -tubulin levels, MT integrity, and cell survival [21,22]. In contrast, mammals possess a TBC-E paralogue, which lacks the CAP-Gly domain. This protein, known as E-like, with  $\sim 30\%$  sequence identity with TBC-E, cannot compensate for loss of TBC-E, and, instead of being involved in tubulin biogenesis, is implicated in degradation [23]. This difference is perhaps a legacy of a more complex and highly regulated tubulin biology in higher eukaryotes, and suggests that a degree of care is required when considering different model systems.

We could not identify any other CAP-Gly domains encoded by the *T. brucei* genome, even in proteins that usually contain such modules, e.g. kinesins. Perhaps, in trypanosomatids, TBC-B is sufficient to provide this interaction with the  $\alpha$ -tubulin tail, or other modules, yet to be discovered, may compensate for this function.

### The missing structural information

As it was not possible to obtain crystallographic data on the linker region between the globular domains (residues 89–153), the CD spectrum of full-length TBC-B was analysed. The spectra indicated a low  $\alpha$ -helical content of  $\sim 10\%$  and a  $\beta$ -strand content of  $\sim 35\%$ . The  $\alpha$ -helical content of the Ubl domain itself is  $\sim 8\%$  of the overall structure, and, together with the prediction of a short helical segment between residues 116 and 123,  $\sim 4\%$  of the sequence, there is excellent agreement with the CD data. Residues with a  $\beta$ -strand conformation constitute nearly 30% of the full-length structure, which is also in good agreement with the spectroscopic data. Both the CD results and the structural analyses also agreed well with the predicted secondary structure, and overall indicate a two-domain structure with a flexible linker.

### Functional implications and concluding remarks

Structures of the Ubl and CAP-Gly domains of *Tb*TBC-B, the first such protist structures, were determined to 2.3-Å and 1.6-Å resolution, respectively. It was not possible to solve the full-length structure, as

the crystals did not diffract sufficiently well, and attempts to extend structural information into the region linking the two domains were also unsuccessful. It was determined that the missing linker region is mostly unstructured, with only a short segment of  $\alpha$ -helix being noted, and potentially flexible.

The fold of the Ubl domain is highly conserved, despite low sequence identity within TBC-B proteins and, indeed, ubiquitin itself. A striking similarity is observed in the spatial location of two functionally important lysines in ubiquitin, and strongly supports a biological function for the Ubl domain, and by implication for TBC-B. Through reversible post-translational modification, TBC-B can influence aspects of tubulin biology in a proteasome-dependent manner. Such a conclusion is consistent with other studies. Yeast two-hybrid studies suggest that proteasome-dependent degradation of human TBC-B is driven by gigaxonin binding to the Ubl domain [24]. In addition, *Saccharomyces cerevisiae* TBC-E interacts with the ubiquitin receptor Rpn10 via the Ubl domain, and subsequently with a ubiquitin ligase complex, providing a route to protein degradation [10]. We note also that Lys34, but not Lys62, of *Tb*TBC-B is conserved in the Ubl domains of TBC-E (data not shown).

In the final stages of MT assembly,  $\alpha$ -tubulin and  $\beta$ -tubulin form a heterodimer with the encouragement of the TBC proteins. A balance has to be struck between the availability of free tubulin, heterodimer assembly, and release of MT structures. This process involves a number of highly abundant proteins, and the presence of Ubl domains offers a means whereby protein folding and the population of complex assemblies can be regulated by protein degradation or recycling. The use of a Ubl protein rather than ubiquitin itself may provide specificity with regard to this aspect of tubulin biology. Cognate activating enzymes might then contribute to the regulation of levels of free tubulin heterodimers, tubulin polymerization, and MT catastrophe, or to avoid miscommunication during a stress response. It will therefore be of great interest to now identify, from a plethora of candidates, the specific ligases and proteases that might participate in the control of MT disassembly/assembly as opposed to other biological functions.

The molecular packing in the crystal structure of the CAP-Gly domain places the C-terminal tail of one CAP-Gly domain in the peptide-binding groove of a symmetry-related molecule, and allows examination of interactions in this binding site. The binding differs slightly from that observed in a CLIP-170 CAP-Gly domain bound to an  $\alpha$ -tubulin tail peptide, owing to a difference in the  $\beta$ 7– $\beta$ 8 tubulin-binding loop of

*Tb*TBC-B. A glycine replaces asparagine at position 3 of the normally conserved GKNDG sequence, and allows the C-terminus of chain A to bind further down in the peptide-binding groove. This appears to be a trypanosomatid-specific feature.

Tubulin and MT assembly are subject to C-terminal post-translational modifications, which provide important tracking positions for a range of MT-binding partners [25]. For example, the C-terminal residue of  $\alpha$ -tubulin is a tyrosine that can be removed and then replaced, and adjacent glutamates can be polyglutamylated or, in some species, polyglycylated. The consensus sequence for recognition by CAP-Gly domains has been deduced as EE[Y/F], and proteolytic removal of the tyrosine prevents such recognition of  $\alpha$ -tubulin [25]. This provides a feature that might be used to regulate subpopulations of MTs. A plausible function of the CAP-Gly domain may therefore be to bind then block post-translational modifications of the C-terminal tyrosine, as the complex with  $\alpha$ -tubulin appears to form even while it is still bound to CCT [8], i.e. while the tyrosine is still present.

We were unable to observe any binding of the CAP-Gly domain with a hexapeptide representing the terminal residues of *T. brucei*  $\alpha$ -tubulin by using ITC. According to previous work, the presence of TBC-E might be required for TBC-B to form a stable complex with  $\alpha$ -tubulin [6]. Further experiments would be required to inform hypotheses involving the need for partner proteins and/or post-translational modifications that allow TBC-B to contribute to MT assembly or disassembly.

## Experimental procedures

### Protein production and purification

The gene encoding full-length *Tb*TBC-B was amplified from genomic DNA (strain 927) with 5'-CATATGTCCGTTGTTAAAGTATCGC-3' and 5'-CTCGAGTTAAAACACCTCCGGGGGAAAGTC-3' as forward and reverse primers, respectively (ThermoFisher Scientific, Waltham, MA, USA). The restriction enzyme sites for *Nde*I and *Xho*I are in bold. The PCR product was ligated into pCR2.1-TOPO with the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA), and then cloned into a modified pET15b (Novagen, Madison, WI, USA) expression vector, which adds an N-terminal His<sub>6</sub> tag and a tobacco etch virus (TEV) protease cleavage site to the product. Sequencing confirmed the identity of the construct, and the vector was heat-shock-transformed into *E. coli* Rosetta (DE3) pLysS cells for expression. DNA encoding the CAP-Gly domain, residues 157–222, was amplified from the vector containing the



*TbTBC-B* gene with 5'-CATATGGCAGAGACAATACA TGTGGGGG-3' and 5'-CTCGAGTTAAACACCTCCG GGGGAAAGTC-3' as forward and reverse primers, respectively (ThermoFisher Scientific). The gene fragment was cloned into the modified pET15b vector as described above, and transformed into *E. coli* Rosetta (DE3) pLysS cells for protein production.

Similar protocols were applied to purify full-length *TbTBC-B* and the CAP-Gly domain. Typically, cells were grown at 37 °C in 1 L of LB medium supplemented with carbenicillin (50 µg·L<sup>-1</sup>) and chloramphenicol (25 µg·L<sup>-1</sup>). Gene expression was induced, at a  $D_{600\text{ nm}}$  of 0.6, by addition of 1 mM isopropyl thio-β-D-galactoside. The culture was incubated for a further 16 h at 22 °C, and cells were then harvested by centrifugation at 3500 *g* for 30 min at 4 °C. The cells were resuspended in a lysis buffer (50 mM Tris/HCl, pH 7.5, 250 mM NaCl, 25 mM imidazole) containing DNase I and EDTA-free protease inhibitors (Roche, Basel, Switzerland), and then lysed with a French press at 16000 p.s.i. The resultant lysate was centrifuged at 35 000 *g* for 30 min at 4 °C, and the supernatant was loaded onto a pre-equilibrated HisTrap HP 5-mL column (GE Healthcare, Milwaukee, WI, USA) precharged with Ni<sup>2+</sup>. A linear gradient of 25 mM to 1 M imidazole was applied to elute the proteins, and the derived fractions were analysed by SDS/PAGE. Samples were pooled and incubated with His-tagged TEV protease at 30 °C for 3 h, and then dialyzed into 50 mM Tris/HCl (pH 7.5) and 250 mM NaCl. The sample was loaded onto a HisTrap HP 5-mL column to remove the His-tagged TEV protease, uncleaved product, and histidine-rich contaminants. A Superdex 20 026/60 size exclusion column (GE Healthcare) equilibrated with 50 mM Tris/HCl (pH 7.5) and 250 mM NaCl was used to further purify the protein. This column had been calibrated with BioRad Gel Filtration standards. Fractions containing the proteins were pooled and concentrated to 10 mg·L<sup>-1</sup> with Amicon Ultra devices (Millipore) for subsequent use. The purity and identity of the proteins were confirmed by MALDI-TOF MS. Yields of ~20 mg·L<sup>-1</sup> cell culture for full-length TBC-B and ~8 mg·L<sup>-1</sup> cell culture for the CAP-Gly domain were obtained.

A series of constructs encoding five polypeptide fragments, covering residues 91–232, 101–232, 113–232, 128–226, and 143–232, were also produced in an attempt to extend structural knowledge between the domains. These polypeptides proved to be either insoluble or failed to crystallize, and so no further details are provided.

### CD of full-length TbTBC-B

CD spectra were recorded on a Jasco J-810 spectropolarimeter. Far-UV CD spectra were obtained with 0.5 mg·mL<sup>-1</sup> protein solutions and a 0.02-cm-pathlength quartz cuvette. Five scans were accumulated and averaged with the following parameters: scan rate, 50 nm·min<sup>-1</sup>; response, 0.5 s; and bandwidth, 1 nm. Protein CD spectra were corrected

by subtracting the appropriate buffer spectrum and correcting for protein concentration. Protein secondary structure estimates were obtained with the CONTIN procedure [26], available from the DICHROWEB server [27].

### Analysis of crystals formed following proteolysis

Crystals of the full-length protein did not diffract beyond ~7-Å resolution (data not shown). Limited proteolysis by addition of chymotrypsin to the crystallization drops [28] was therefore tested in the search for ordered crystals. Crystals were observed in a number of conditions, one of which was successfully optimized, as detailed below. These crystals were harvested, washed in the reservoir solution, dissolved in ddH<sub>2</sub>O, and then submitted for MALDI-TOF MS analysis. This gave a molecular mass of 12 069 Da for the polypeptide. The fragment was isolated by SDS/PAGE and then transferred onto a poly(vinylidene difluoride) membrane. The band was then submitted for N-terminal Edman sequencing, and the sequence GHMSVVKV was identified. This corresponds to the N-terminus of *TbTBC-B*, with the first two residues being remnants of the TEV cleavage site after treatment to remove the His<sub>6</sub> tag. These data indicated that the Ubl domain had crystallized.

### Isolation of products after chymotrypsin treatment

Full-length *TbTBC-B* (50 mg) was incubated with chymotrypsin (0.02 mg) overnight, and then dialyzed into 50 mM Tris/HCl (pH 7.5) and 50 mM NaCl. The mixture was loaded onto a HiTrap Q HP 5-mL column, and eluted with a linear gradient of 50 mM to 1 M NaCl. The products of the cleavage eluted as two peaks, and were analysed by SDS/PAGE. Sample 1 eluted at 50 mM NaCl, and sample 2 at 220 mM NaCl. MALDI-TOF MS of sample 1, the Ubl domain, gave a mass of 12 002 Da, and MALDI-TOF MS of sample 2 gave a mass of 12 116 Da. Both protein fragment samples were concentrated to 5 mg·mL<sup>-1</sup>.

### Selenomethionine (SeMet) derivative of the Ubl domain

A SeMet-substituted Ubl domain was obtained from material generated as described above, and with incorporation achieved following metabolic inhibition [29]. Purification of the protein yielded ~2 mg·L<sup>-1</sup> of cell culture. Analysis by MALDI-TOF MS indicated full incorporation of four SeMet residues.

### Crystallization and data collection

Initial crystallization conditions were established by sitting drop vapour diffusion with a Phoenix Liquid Handling

**Table 1.** TbTBC-B domain crystallographic statistics.

|                                                                               | Ubl domain                               | CAP-Gly domain                                        |
|-------------------------------------------------------------------------------|------------------------------------------|-------------------------------------------------------|
| Space group                                                                   | <i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2 | <i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Unit cell lengths (Å)                                                         | 50.9, 50.9, 77.2                         | 32.7, 55.7, 80.4                                      |
| Resolution range (Å)                                                          | 36.00–2.35                               | 40.18–1.59                                            |
| Completeness (%)                                                              | 100 (99.8) <sup>a</sup>                  | 100 (99.97)                                           |
| <I/σ(I)>                                                                      | 15.4 (3.3)                               | 28.2 (9.6)                                            |
| Reflections measured/<br>unique                                               | 58 686/4631                              | 282 648/20 384                                        |
| Redundancy                                                                    | 12.7 (9.4)                               | 13.9 (14.1)                                           |
| Anomalous redundancy:                                                         | 7.1 (5.0)                                | NA                                                    |
| <i>R</i> <sub>merge</sub> <sup>b</sup>                                        | 0.12 (0.64)                              | 0.07 (0.28)                                           |
| <i>R</i> <sub>work</sub> <sup>c</sup> / <i>R</i> <sub>free</sub> <sup>d</sup> | 0.16/0.21                                | 0.20/0.24                                             |
| Protein residues (chain A/B)                                                  | 89                                       | 79/78                                                 |
| Ligands (number)                                                              | Ethylene glycol (1)                      | Formate (6)                                           |
| Rmsd from ideal geometry                                                      |                                          |                                                       |
| Bond lengths (Å)/angles (°)                                                   | 0.015, 1.53                              | 0.01, 1.44                                            |
| Thermal parameters ( <i>B</i> , Å <sup>2</sup> )                              |                                          |                                                       |
| Wilson <i>B</i>                                                               | 42.6                                     | 14.6                                                  |
| Mean <i>B</i> (all atoms)                                                     | 20.1                                     | 14.8                                                  |
| Protein atoms (chain A/B)                                                     | 18.0                                     | 12.3/12.5                                             |
| Ligands                                                                       | 58.9                                     | 21.0                                                  |
| Water molecules                                                               | 42.4                                     | 24.8                                                  |
| Ramachandran plot                                                             |                                          |                                                       |
| Favored/allowed (%)                                                           | 97.6/2.4                                 | 98/2                                                  |

NA, not analysed.

<sup>a</sup> Values in parentheses refer to the highest-resolution bin.<sup>b</sup>  $R_{\text{merge}} = \sum h \sum i |I(h, i) - \langle I(h) \rangle| / \sum h \sum i I(h, i)$ .<sup>c</sup>  $R_{\text{work}} = \sum h k l |F_o - I F_c| / \sum h k l F_o$ , where  $F_o$  is the observed structure factor amplitude and  $F_c$  is the structure-factor amplitude calculated from the model.<sup>d</sup>  $R_{\text{free}}$  is the same as  $R_{\text{work}}$  except that it was only calculated using a subset, 5%, of the data that are not included in any refinement calculations.

System (Rigaku; Art Robins Instruments, Mountain View, CA, USA) and commercially available screens. Several conditions were optimized for hanging drop vapour diffusion, with drops of 1 µL of protein solution and 1 µL of reservoir.

Rectangular crystals (dimensions 0.3 × 0.1 × 0.1 mm) of the Ubl domain, which had been isolated after proteolysis of the full-length protein, were obtained with a reservoir of 0.1 M Hepes (pH 7.5), 1% poly(ethylene glycol) 400, and 2 M ammonium sulfate, and a protein concentration of 2.5 mg·mL<sup>-1</sup>. Crystals of the CAP-Gly domain, derived from expression of a gene fragment, were rod-like, with dimensions of 1 × 0.05 × 0.05 mm, and were obtained with reservoir conditions of 0.2 M potassium formate, 30% poly(ethylene glycol) 3350 and a 7.5 mg·mL<sup>-1</sup> protein solution in 50 mM Tris/HCl (pH 7.5) and 250 mM NaBr.

Crystals were transferred to a solution containing a mixture of their reservoir solution and 40% poly(ethylene glycol) 400 for ~15 s before being cooled to -173 °C in a stream of nitrogen. The crystals were characterized in-house with a Micromax HF007 copper-rotating anode

X-ray generator equipped with an R-axis IV<sup>++</sup> dual image plate detector. Subsequently, data were collected with ADSC charge-coupled device detectors at the Diamond light source beamline I03 for the Ubl domain and at beamline I02 for the CAP-Gly domain.

Data from the Ubl domain crystal were processed with XDS [30] and SCALA [31]. Initial phases were obtained with SeMet single-wavelength anomalous dispersion methods [14], with data measured at the experimentally determined *f'* maximum wavelength ( $\lambda = 0.98$  Å) in the CCP4 pipeline [32] with CRANK [33]. A figure-of-merit of 0.23 was obtained from BP3 [34], and this increased to 0.67 after density modification and solvent flattening in SOLOMON [35]. This phase set was used for initial model building with BUCCANEER [36].

Data from the CAP-Gly domain crystal were scaled and processed with IMOSFLM [37] and SCALA [31]. The PHRE server [38] identified the NMR structure of the *M. musculus* CAP-Gly domain of TBC-B (PDB code [1WHG](#)), sharing a sequence identity of ~55%, as a suitable model for molecular replacement. A poly-Ala model was prepared with CHAINSAW [39], and the positions of two molecules were then identified with PHASER [40]. Rigid body refinement gave an initial *R*-factor of 49% to 1.6-Å resolution.

Both structures were refined with rounds of map inspection and model manipulation with COOT [41] and refinement calculations with REFMAC [42]. When the protein models were nearly complete, waters and ligands (ethylene glycol in the Ubl domain and formate in the CAP-Gly domain) were added, together with multiple conformers for several side chains. The geometric quality of the models was assessed with MOLPROBITY [43]. Figures were created with PYMOL [44], and solvent accessibility analyses were performed with AREAIMOL [32]. Crystallographic statistics are presented in Table 1. Coordinates and structure factors are deposited with the PDB under accession codes [4B6W](#) for the Ubl domain and [4B6M](#) for the Cap-Gly domain.

## Bioinformatic analyses

Secondary structure and disorder predictions were obtained from PHYRE [38] and PSIPRED [45]. Structural comparisons and homologues were identified with DALI [46]. Conserved residues within the CAP-Gly domain were identified with CONSURF [47]. The UniProt reference cluster 90 ([UNI-REF90](#)) [48] database was searched, and sequences were aligned with MAFFT [49]. Overall, 642 sequences were identified with PSI BLAST [50], on the basis of sequence identity in the range 35–95%, 417 of which were unique. A subset of sequences, 150, the default value in CONSURF, with identity > 50% were used for comparisons. However, there were not enough homologues with identities of > 30% of the Ubl domain for CONSURF to be used, so 31 sequences with identities in the range 31–38%, were aligned by the use of MUSCLE [51], and residues with > 75% conservation were annotated.

## ITC details

ITC experiments were performed at 25 °C with a VP-ITC calorimeter (MicroCal, Northampton, MA, USA). The 1.4-mL sample cell and 300- $\mu$ L syringe were filled with TBC-B CAP-Gly domain and peptide EDVEEY in a buffer of 50 mM Tris/HCl (pH 7.5) and 250 mM NaCl. Typically, 8  $\mu$ L of peptide was injected from the stirred syringe (305 r.p.m.) 30 times into the sample cell. The peptides were dissolved in the minimum amount of the buffer, and then diluted to the appropriate concentration just before use. Four experiments were carried out with CAP-Gly/hexapeptide concentrations of 1 : 10, 10 : 100, 7 : 500 and 45 : 500  $\mu$ M. No binding events were observed.

## Acknowledgements

This work was supported by BBSRC (PhD studentship to J. Fleming) and The Wellcome Trust (grant numbers WT082596, WT083481, and WT094090). We thank Diamond for synchrotron beam time and staff support, and our colleague D. Norman for discussions and help.

## References

- Szymanski D (2002) Tubulin folding cofactors: half a dozen for a dimer. *Curr Biol* **12**, R767–R769.
- Lundin VF, Leroux MR & Stirling PC (2010) Quality control of cytoskeletal proteins and human disease. *Trends Biochem Sci* **35**, 288–297.
- Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL & Cowan NJ (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**, 863–873.
- Muñoz IG, Yébenes H, Zhou M, Mesa P, Serna M, Park AY, Bragado-Nilsson E, Beloso A, de Cárcer G, Malumbres M *et al.* (2011) Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat Struct Mol Biol* **18**, 14–19.
- Grynberg M, Jaroszewski L & Godzik A (2003) Domain analysis of the tubulin cofactor system: a model for tubulin folding and dimerization. *BMC Bioinformatics* **10**, 46.
- Kortazar D, Fanarraga ML, Carranza G, Bellido J, Villegas JC, Avila J & Zabala JC (2007) Role of cofactors B (TBCB) and E (TBCE) in tubulin heterodimer dissociation. *Exp Cell Res* **313**, 425–436.
- Lopez-Fanarraga M, Carranza G, Bellido J, Kortazar D, Villegas JC & Zabala JC (2007) Tubulin cofactor B plays a role in the neuronal growth cone. *J Neurochem* **100**, 1680–1687.
- Carranza G, Castaño R, Fanarraga ML, Villegas JC, Gonçalves J, Soares H, Avila J, Marenchino M, Campos-Olivas R, Montoya G *et al.* (2013) Autoinhibition of TBCB regulates EB1-mediated microtubule dynamics. *Cell Mol Life Sci* **70**, 357–371.
- Baffet AD, Benoit B, Januschke J, Audo J, Gourhand V, Roth S & Guichet A (2012) *Drosophila* tubulin-binding cofactor B is required for microtubule network formation and for cell polarity. *Mol Biol Cell* **23**, 3591–3601.
- Voloshin O, Gocheva Y, Gutnick M, Movshovich N, Bakhrat A, Baranes-Bachar K, Bar-Zvi D, Parvari R, Gheber L & Raveh D (2010) Tubulin chaperone E binds microtubules and proteasomes and protects against misfolded protein stress. *Cell Mol Life Sci* **67**, 2025–2038.
- Lytle BL (2004) Solution structure of a ubiquitin-like domain from tubulin-binding cofactor B. *J Biol Chem* **279**, 46787–46793.
- Li S, Finley J, Liu ZJ, Qiu SH, Chen H, Luan CH, Carson M, Tsao J, Johnson D, Lin G *et al.* (2002) Crystal structure of the cytoskeleton-associated protein glycine-rich (CAP-Gly) domain. *J Biol Chem* **277**, 48596–48601.
- Ralston KS, Kabututu ZP, Melehan JH, Oberholzer M & Hill KL (2009) The *Trypanosoma brucei* flagellum: moving parasites in new directions. *Annu Rev Microbiol* **63**, 335–362.
- Micossi E, Hunter WN & Leonard GA (2002) *De novo* phasing of two crystal forms of trypanothione II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures. *Acta Crystallogr D Biol Crystallogr* **58**, 21–28.
- Ikeda F & Dikic I (2008) Atypical ubiquitin chains: new molecular signals. 'Protein modifications: beyond the usual suspects' review series. *EMBO Rep* **9**, 536–542.
- Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503–533.
- Weisbrich A, Honnappa S, Jaussi R, Okhrimenko O, Frey D, Jelesarov I, Akhmanova A & Steinmetz MO (2007) Structure–function relationship of CAP-Gly domains. *Nat Struct Mol Biol* **14**, 959–967.
- Steinmetz MO & Akhmanova A (2008) Capturing protein tails by CAP-Gly domains. *Trends Biochem Sci* **33**, 535–545.
- Mishima M, Maesaki R, Kasa M, Watanabe T, Fukata M, Kaibuchi K & Hakoshima T (2007) Structural basis for tubulin recognition by cytoplasmic linker protein 170 and its autoinhibition. *Proc Natl Acad Sci USA* **104**, 10346–10351.
- Honnappa S, Okhrimenko O, Jaussi R, Jawhari H, Jelesarov I, Winkler FK & Steinmetz MO (2006) Key interaction modes of dynamic +TIP networks. *Mol Cell* **23**, 663–671.
- Radcliffe PA, Garcia MA & Toda T (2000) The cofactor-dependent pathways for alpha- and

- beta-tubulins in microtubule biogenesis are functionally different in fission yeast. *Genetics* **156**, 93–103.
- 22 Peris L, Thery M, Fauré J, Saoudi Y, Lafanechère L, Chilton JK, Gordon-Weeks P, Galjart N, Bornens M, Wordeman L *et al.* (2006) Tubulin tyrosination is a major factor affecting the recruitment of CAP-Gly proteins at microtubule plus ends. *J Cell Biol* **174**, 839–849.
  - 23 Bartolini F, Tian G, Piehl M, Cassimeris L, Lewis SA & Cowan NJ (2005) Identification of a novel tubulin-destabilizing protein related to the chaperone cofactor E. *J Cell Sci* **118**, 1197–1207.
  - 24 Wang W, Ding J, Allen E, Zhu P, Zhang L, Vogel H & Yang Y (2005) Gigaxonin interacts with tubulin folding cofactor B and controls its degradation through the ubiquitin–proteasome pathway. *Curr Biol* **15**, 2050–2055.
  - 25 Janke C & Bulinski JC (2011) Post-translational regulation of the microtubule cytoskeleton: mechanisms and functions. *Nat Rev Mol Cell Biol* **12**, 773–786.
  - 26 Provencher S & Gloeckner J (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37.
  - 27 Whitmore L & Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, 668–673.
  - 28 Dong A, Xu X & Edwards AM (2007) *In situ* proteolysis for protein crystallization and structure determination. *Nat Methods* **4**, 1019–1021.
  - 29 Doublé S (1997) Preparation of selenomethionine proteins for phase determination. *Methods Enzymol* **276**, 523–530.
  - 30 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
  - 31 Evans P (2005) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72–82.
  - 32 Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763.
  - 33 de Graaff R, Abrahams J & Pannu N (2004) Crank: new methods for automated macromolecular crystal structure solution. *Structure* **12**, 1753–1761.
  - 34 Storoni L & Read R (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr D Biol Crystallogr* **60**, 1220–1228.
  - 35 Abrahams JP & Leslie AGW (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**, 30–42.
  - 36 Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002–1011.
  - 37 Battye TGG, Kontogiannis L, Johnson O, Powell HR & Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271–281.
  - 38 Kelley LA & Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.
  - 39 Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* **41**, 641–643.
  - 40 McCoy A, Grosse-Kunstleve R, Adams P, Winn M, Storoni L & Read R (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
  - 41 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
  - 42 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
  - 43 Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, 375–383.
  - 44 DeLano W (2002) Pymol: an open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography.
  - 45 Buchan DWA, Ward SM, Lobley AE, Nugent TCO, Bryson K & Jones DT (2010) Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38**, 563–568.
  - 46 Holm L, Kääriäinen S, Wilton C & Plewczynski D (2006) Using Dali for structural comparison of proteins. *Curr Protoc Bioinformatics* 14:5.5.1–5.5.24.
  - 47 Ashkenazy H, Erez E, Martz E, Pupko T & Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529–W533.
  - 48 Suzek BE, Huang H, McGarvey P, Mazumder R & Wu CH (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288.
  - 49 Katoh K, Misawa K, Kuma K-I & Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066.
  - 50 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
  - 51 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.